J. Phys. A: Math. Gen. **30** (1997) 7677–7693. Printed in the UK PII: S0305-4470(97)84026-7

Variational method and duality in the 2D square Potts model

L Angelini†, M Pellicoro‡, I Sardella§ and M Villani‖
Dipartimento di Fisica dell'Università di Bari, 70126 Bari, Italy
and
Istituto Nazionale di Fisica Nucleare, Sezione di Bari, 70126 Bari, Italy

Received 7 May 1997, in final form 28 July 1997

Abstract. The ferromagnetic q-states Potts model on a square lattice is analysed, for $q > 4$, through an elaborate version of the operatorial variational method. In the variational approach proposed in this paper, the duality relations are exactly satisfied, involving at a more fundamental level, a duality relationship between variational parameters. Besides some exact predictions, the approach is very effective in the numerical estimates over a wide range of temperature and can be systematically improved.

1. Introduction

We consider the ferromagnetic q-state Potts model [1] on a square lattice, limiting ourselves to the isotropic, nearest-neighbour (NN) case, with no external field. The Hamiltonian of the model is given by

$$\frac{H}{kT} = -K \sum_{\langle ij \rangle} \delta_{s_i s_j} \tag{1}$$

where $s_i = 1, 2, \ldots, q$, i denotes the lattice sites, $\delta_{s_i s_j}$ is the Kronecker symbol and $K = J/kT$, with $J > 0$. We suppose that the lattice is made of m rows and n columns, and that free boundary conditions are applied. Our interest will be in the limits $m, n \to \infty$.

For $q > 4$ the model has a temperature-driven first-order phase transition at $K_t = K_t(q) = \ln(1 + \sqrt{q})$. At K_t the free energy f, the internal energies u_o and u_d of the ordered and disordered phase, and the latent heat are known exactly [1, 2]. Still at K_t, more recently an analytic formula for the correlation length has been also derived [3, 4].

Due to these exact results, the 2D Potts model has an important role in the study of first-order phase transitions [5]. As a matter of fact its properties have been studied through several methods, such as analysis of high- or low-temperature series expansions [6], resummation of large q expansions [7–9] and Monte Carlo simulations [10–14]. As a rule, the discontinuous character of the first-order phase transitions makes the numerical calculations in these cases more uncertain, with respect to analogous calculations in the case of second-order phase transitions. Moreover we have that some aspects of the discontinuous transitions are not completely understood from a general, theoretical point of view.

† E-mail address: angelini@@bari.infn.it
‡ E-mail address: pellicoro@@bari.infn.it
§ E-mail address: sardella@@bari.infn.it
‖ E-mail address: villani@@bari.infn.it

0305-4470/97/227677+17$19.50

On the other hand an accurate description of the Potts model transition, when the number of components q is very large, is provided by the mean-field theory [15]. Indeed, it has been proved that mean-field theory is exact in the limit $q \to +\infty$ [16]. So it suggests itself that a reliable quantitative analysis of the properties of our model, at least for $q > 4$, could also be obtained directly from a proper extension of this theory.

As a matter of fact, in this paper it is shown that considerable and systematic improvements of the mean-field theory can be obtained in the framework of the transfer-matrix approach, through an elaborate version of the standard variational method for Hermitian operators. This method is alternative to the variational Gibbs principle [16, 17], which provides the analytical basis of the mean-field procedure or, more generally, of the cluster variational method [18].

The formal aspects of our approach are described in the next four sections. We will limit ourselves to $q > 4$. In the last section we present the results of some numerical calculations and add some concluding remarks.

2. Operatorial variational method

Let $Z_{m,n}(K)$ be the partition function of our model. We are interested in the free energy per site, $f(T)$, given by

$$f(T) = -kT\phi(K)$$

where

$$\phi(K) = \lim_{m\to\infty}\lim_{n\to\infty}\frac{1}{nm}\log Z_{m,n}(K). \tag{2}$$

Now, if $\sigma = (s_1, s_2, \ldots, s_m)$ describes the spin configuration on a column, we can write $Z_{m,n}$ in the form

$$Z_{m,n} = \sum_{\sigma_1,\sigma_2,\ldots,\sigma_n} \chi(\sigma_1)L(\sigma_1,\sigma_2)L(\sigma_2,\sigma_3)\ldots L(\sigma_{n-1},\sigma_n)\chi(\sigma_n) = (\chi, L^{n-1}\chi) \tag{3}$$

where

$$\chi(\sigma) = \mathrm{e}^{\frac{K}{2}\sum_{i=1}^{m-1}\delta_{s_i,s_{i+1}}} \tag{4}$$

and

$$L(\sigma,\sigma') = \chi(\sigma)\mathrm{e}^{K\sum_{i=1}^{m}\delta_{s_i,s_i'}}\chi(\sigma') \tag{5}$$

is the symmetrized transfer matrix of our model.

The matrix $L(\sigma,\sigma')$ is Hermitian. Therefore we can apply the spectral theorem and take account of the extremal properties of its eigenvalues. We also note that, for our $L(\sigma,\sigma')$, the Perron–Frobenius (PF) theorem is valid [19]. Furthermore we have that $L(\sigma,\sigma')$ is positive definite, that is, for any vector $\xi(\sigma)$, we obtain

$$(\xi, L\xi) = \sum_{\sigma,\sigma'}\bar{\xi}(\sigma)L(\sigma,\sigma')\xi(\sigma') \geqslant 0. \tag{6}$$

In fact we can write

$$(\xi, L\xi) = \sum_{\sigma,\sigma'}\bar{f}(\sigma)\prod_{i=1}^{m}(1 + (\mathrm{e}^K - 1)\delta_{s_i,s_i'})f(\sigma')$$

where

$$f(\sigma) = \xi(\sigma)\chi(\sigma).$$

By expanding the product, we obtain a sum of 2^m terms, each giving a non-negative contribution, due to $e^K - 1 > 0$.

Let $\lambda_1(K; m)$ be the highest eigenvalue of $L(\sigma, \sigma')$. From (2) and (3) it follows that [19]:

$$\phi(K) = \lim_{m\to\infty} \frac{1}{m} \log \lambda_1(K, m). \tag{7}$$

Now, by considering the functional given by the Rayleigh–Ritz (RR) quotient

$$F[\xi] = \frac{(\xi, L\xi)}{(\xi, \xi)} \tag{8}$$

we have

$$\sup_{\xi} F[\xi] = \lambda_1(K; m)$$

the sup being attained for $\xi(\sigma) = \psi_1(\sigma)$, where $\psi_1(\sigma)$ is the eigenvector of $L(\sigma, \sigma')$ corresponding to $\lambda_1(K; m)$. So, for any vector $\xi(\sigma)$,

$$\lim_{m\to\infty} \frac{1}{m} \log \frac{(\xi, L\xi)}{(\xi, \xi)} \leqslant \phi(K). \tag{9}$$

This inequality is the starting point of the operatorial variational method in statistical mechanics [20–24]. In this approach, the principal (or ground state) eigenvector $\psi_1(\sigma)$ is usually described through an ansatz $\tilde{\psi}(\sigma; \alpha, \beta, \ldots)$, having a definite formal structure, which contains one or several unknown parameters which do not depend on m. Then, within the formal structure of the considered ansatz, the best estimate of $f(T)$ is obtained by fixing the parameters through

$$\sup_{\alpha,\beta,\ldots} \lim_{m\to\infty} \frac{1}{m} \log \frac{(\tilde{\psi}, L\tilde{\psi})}{(\tilde{\psi}, \tilde{\psi})}. \tag{10}$$

As a matter of fact, this approach, which is at the origin of the concept of the transfer matrix, is not really adequate to analyse the properties of a critical point. However, the situation can be different in the case of first-order transitions.

We note that the above approach can be realized in several ways, depending on the structure of the chosen ansatz and on the way the sup in (10) is handled. One elegant and efficient particular structure has been introduced by Baxter [2, 22, 23], which allows us to obtain explicitly the stationary points of the RR quotient through the solutions of a set of nonlinear matrix equations, which are treated numerically. This structure is improved by changing only the order of the unknown matrices. However other efficient structures of the ansatz can also be introduced [21] on the basis of physical and mathematical arguments; in these cases the sup (10) is handled directly by numerical procedures.

In the following, fixing attention on the 2D square Potts model, we will discuss a further elaboration of the operatorial variational method. Our aim, compared with previous realizations, is to see if the variational method, due to its non-perturbative nature, can be adequately realized in such a way that some general non-perturbative aspects, such as duality, can be obtained exactly. So, we come back to the problem of the structure of the variational ansatz, by adopting a more flexible point of view.

First of all, it is useful to bear in mind the meaning and the role of $\psi_1(\sigma)$. In our lattice made of m rows, let us consider a strip Σ_i of i adjacent columns, having n_r columns at its right and n_l columns at its left. The reduced probability distribution (RPD)

$P_i = P_i(\sigma_1, \sigma_2, \ldots, \sigma_i)$ of a spin configuration on Σ_i, obtained by averaging over all the configurations at the right and the left of Σ_i, is given, in the limit $n_r, n_l \to +\infty$, by

$$P_i(\sigma_1, \sigma_2, \ldots, \sigma_i) = \frac{\psi_1(\sigma_1)L(\sigma_1, \sigma_2)L(\sigma_2, \sigma_3)\ldots L(\sigma_{i-1}, \sigma_i)\psi_1(\sigma_i)}{(\psi_1, L^{i-1}\psi_1)} \tag{11}$$

for $i \geqslant 2$, while in the case of one column, we have the RPD $P_1(\sigma)$ through

$$P_1(\sigma) = \frac{\psi_1^2(\sigma)}{(\psi_1, \psi_1)}. \tag{12}$$

So, an ansatz $\tilde{\psi}(\sigma)$ for $\psi_1(\sigma)$ has the role of a parametric description of the above averaging process leading to the RPD P_i.

Now, very reliable and simple descriptions, having, however, a different formal content, can be obtained by considering separately the high- and low-temperature regions. This fact drives us to introduce for $\tilde{\psi}(\sigma)$ two formal structures, generally different, which we call $\tilde{\psi}_{\rm d}(\sigma)$ and $\tilde{\psi}_{\rm o}(\sigma)$ and to consider $\tilde{\psi}(\sigma)$ as made of two *charts* $(\tilde{\psi}_{\rm d}(\sigma), \tilde{\psi}_{\rm o}(\sigma))$, associated with the ordered and disordered phase, which should match, if possible, at the transition point $K_{\rm t}$. The choice of the appropriate *chart* is demanded by the variational method. This point of view is a slight generalization of the traditional approach, where a fixed formal structure for $\tilde{\psi}(\sigma)$ is considered, and only the parameters are varied. However, it leads quite naturally to variational approximations to $\phi(K)$, in which duality [2, 25] is exactly satisfied.

3. One-parameter ansatz

In this section we introduce and discuss some one-parameter ansatz of which we will make use in the variational approach based on equation (10).

Let us fix attention first on the disordered phase. For large T, due to the smallness of the correlation length, we expect that

$$P_1(\sigma) \simeq \frac{{\rm e}^{K\sum_{i=1}^{m-1}\delta_{s_i,s_{i+1}}}}{(\chi, \chi)}. \tag{13}$$

Then, with regard to the disordered phase, we introduce the simplest one-parameter ansatz for $\psi_1(\sigma)$ in the form

$$\tilde{\psi}_{\rm d}^{(1)}(\sigma) = {\rm e}^{\frac{A(K)}{2}\sum_{i=1}^{m-1}\delta_{s_i,s_{i+1}}} \qquad (A(K) > 0) \tag{14}$$

that is, for K belonging to the disordered region, the result of the averaging process at the left and the right of a column, is represented simply by a renormalized ferromagnetic coupling $A(K)$ between NN spins, such that $A(K) \simeq K$ for small K.

This simplest ansatz can be improved in several ways. Since the averaging process produces new interactions among the spins of a column, we need, in principle, other parameters related to these new couplings. For example, a first correction to $\tilde{\psi}_{\rm d}^{(1)}(\sigma)$ can be considered by also taking into account second-neighbour interactions. However, it is possible to stay with a one-parameter ansatz and, at the same time, to improve $\tilde{\psi}_{\rm d}^{(1)}(\sigma)$, by applying the transfer matrix L to $\tilde{\psi}_{\rm d}^{(1)}(\sigma)$. As a matter of fact, if L is applied repeatedly to $\tilde{\psi}_{\rm d}^{(1)}(\sigma)$, we reach at the end the principal eigenvector. Since L is positive definite, we can consider, more generally, the action of $L^{\frac{1}{2}}$. So, due to the simple structure of $\tilde{\psi}_{\rm d}^{(1)}(\sigma)$, we will fix our attention on the sequence of one-parameter ansatz

$$\tilde{\psi}_{\rm d}^{(1)}(\sigma), \tilde{\psi}_{\rm d}^{(2)}(\sigma) = L^{\frac{1}{2}}\tilde{\psi}_{\rm d}^{(1)}(\sigma), \ldots, \tilde{\psi}_{\rm d}^{(\nu)}(\sigma) = L^{\frac{\nu}{2}}\tilde{\psi}_{\rm d}^{(1)}(\sigma). \tag{15}$$

The reliability of the one-approximate description of $\psi_1(\sigma)$ can be tested by considering the next approximation in the sequence.

Then we are led to the RR quotients

$$\rho_{\rm d}^{(1)}(A,K)=\frac{(\tilde\psi_{\rm d}^{(1)},L\tilde\psi_{\rm d}^{(1)})}{(\tilde\psi_{\rm d}^{(1)},\tilde\psi_{\rm d}^{(1)})}\qquad \rho_{\rm d}^{(2)}(A,K)=\frac{(\tilde\psi_{\rm d}^{(2)},L\tilde\psi_{\rm d}^{(2)})}{(\tilde\psi_{\rm d}^{(2)},\tilde\psi_{\rm d}^{(2)})},\ldots \tag{16}$$

which can also be written in the form

$$\rho_{\rm d}^{(\nu)}(A,K)=\frac{(\tilde\psi_{\rm d}^{(1)},L^{\nu}\tilde\psi_{\rm d}^{(1)})}{(\tilde\psi_{\rm d}^{(1)},L^{\nu-1}\tilde\psi_{\rm d}^{(1)})}. \tag{17}$$

We note that

$$Z_{\rm d}^{(\nu)}(A,K)=(\tilde\psi_{\rm d}^{(1)},L^{\nu-1}\tilde\psi_{\rm d}^{(1)}) \tag{18}$$

is, for $\nu>1$, the partition function of a Potts(ν; A) model on a strip Σ_ν of m rows and ν columns, having the same couplings as the original model (1), except on the first and last column, where there is vertical NN coupling $(A(K)+K)/2$. This model provides an approximate description $\tilde P_{\rm d}^{(\nu)}$ of the RPD $P_\nu(\sigma_1,\sigma_2,\ldots,\sigma_\nu)$.

We also remark that the Potts(ν; A) is associated with a $q^\nu\times q^\nu$ transfer matrix $\ell_\nu(A,K)$, connecting two adjacent rows of Σ_ν. We will call $\lambda_1^{(\nu)}(A,K)$ the highest eigenvalue of $\ell_\nu(A,K)$.

In each RR quotient we can fix the parameter $A(K)$, by considering the variational estimate (10). We obtain then a sequence of estimates $\phi_{\rm d}^{(\nu)}(K)$ of $\phi(K)$, given by

$$\begin{aligned}\phi_{\rm d}^{(\nu)}(K)&=\sup_A\lim_{m\to\infty}\frac{1}{m}\log\frac{Z_{\rm d}^{(\nu+1)}(A,K)}{Z_{\rm d}^{(\nu)}(A,K)}\\&=\sup_A\log\frac{\lambda_1^{(\nu+1)}(A,K)}{\lambda_1^{(\nu)}(A,K)}\equiv\sup_A\phi_{\rm d}^{(\nu)}(A,K)\qquad(\nu=1,2,\ldots)\end{aligned} \tag{19}$$

where the sup is taken over values of A which give ferromagnetic couplings, that is $A\geqslant 0$ for $\nu=1$, $A\geqslant -K$ for $\nu=2,3,\ldots$.

It is useful to point out that, for $0\leqslant A$, $K<+\infty$, the PF theorem is valid for the *reduced* transfer matrices of finite order $\ell_\nu(A,K)$, so that, in this range of parameters, the Potts(ν; A) model has a unique *ground state* (as it happens in the thermodynamic limit for the original model) in the disordered phase.

Now, let us consider the ordered phase. For finite m, even if large, $\psi_1(\sigma)$ is a well defined, unique, vector; furthermore $P_1(\sigma)$ has the same symmetries of the Hamiltonian (1). However, in the limit $m\to\infty$, we have several phenomena as asymptotic degeneracy [2] of the highest eigenvalue of L (related to the occurrence of q ordered phases), macroscopic instabilities with respect to boundary perturbations or the existence of *spontaneous* long-range order. So, an ansatz $\tilde\psi_{\rm o}(\sigma)$ will be appropriate if these aspects can be correctly predicted. In particular, if we introduce *reduced* transfer matrices of finite order, as we did in the disordered case, we have to elude the PF theorem.

We show that the above requirements can be satisfied through an approximate description not of the result of the averaging process leading to $\psi_1(\sigma)$, but the process itself.

Let us consider a column Σ. We replace the n_r columns, with $n_r\to\infty$, at the right of Σ, with one column Σ'' of Potts spins $\sigma_1'',\sigma_2'',\ldots,\sigma_m''$, having an effective horizontal coupling $H(H>0)$ with the spins of Σ. Furthermore, adjacent spins σ_i'' are constrained to be in the same state, this last condition having the role to describe the ordered phase. The averaging process at the right of Σ is then simulated by the summation over all configurations on Σ''.

The same procedure is applied at the left of Σ, by introducing a column Σ'. So, we obtain the distribution

$$\tilde{P}_0(\sigma) = \frac{1}{\Gamma} \sum_{\sigma'} \sum_{\sigma''} \Pi(\sigma', \sigma)\Pi(\sigma'', \sigma) e^{K \sum_{i=1}^{m-1} \delta_{s_i, s_{i+1}}} \tag{20}$$

where Γ is a normalization constant and

$$\Pi(\sigma', \sigma) = e^{H \sum_{i=1}^{m} \delta_{s'_i, s_i}} \prod_{j=1}^{m-1} \delta_{s'_j, s'_{j+1}}. \tag{21}$$

We are then led to the one-parameter ansatz $\tilde{\psi}_0^{(1)}(\sigma)$ in the ordered phase, given by

$$\tilde{\psi}_0^{(1)}(\sigma) = \left(\sum_{\sigma'} \Pi(\sigma', \sigma) \right) e^{\frac{K}{2} \sum_{i=1}^{m-1} \delta_{s_i, s_{i+1}}}. \tag{22}$$

We note that $\tilde{P}_0(\sigma)$ is a RPD obtained from the distribution

$$\tilde{P}_0(\sigma', \sigma, \sigma'') = \frac{1}{\Gamma} \Pi(\sigma', \sigma)\Pi(\sigma'', \sigma) e^{K \sum_{i=1}^{m-1} \delta_{s_i, s_{i+1}}} \tag{23}$$

involving the strip $\Sigma' \cup \Sigma \cup \Sigma''$. Now, $\tilde{P}_0(\sigma', \sigma, \sigma'')$ is associated with a $q^3 \times q^3$ transfer matrix $\ell = \ell(s'_i, s_i, s''_i | s'_{i+1}, s_{i+1}, s''_{i+1})$, having many matrix elements equal to zero, due to the factors $\prod_{i=1}^{m-1} \delta_{s'_i, s'_{i+1}}$, $\prod_{i=1}^{m-1} \delta_{s''_i, s''_{i+1}}$. This allows us to elude the PF theorem. In fact, the highest eigenvalue of ℓ is q times degenerate; we have that ℓ is the direct sum of the q blocks made by the same $q \times q$ matrix,

$$\ell(\mu) = \ell(s_i, s_{i+1}; \mu) = \ell(\mu, s_i, \mu | \mu, s_{i+1}, \mu) \qquad (\mu = 1, 2, \ldots, q)$$

and of the $q(q-1)$ blocks

$$\ell(\mu, \mu') = \ell(s_i, s_{i+1}; \mu, \mu') = \ell(\mu, s_i, \mu' | \mu, s_{i+1}, \mu') \qquad (\mu, \mu' = 1, 2, \ldots, q, \ \mu \neq \mu')$$

made also by $q \times q$ coincident matrices, which differ from the previous ones. The highest eigenvalue of ℓ is contained in the blocks $\ell(\mu)$, which are associated with the q ordered phases, while $\ell(\mu, \mu')$ are related to the occurrence of a surface tension between different phases.

The problem of selecting a particular $\ell(\mu)$ is physically equivalent to applying an external field h_μ to the spin state μ, on the boundaries Σ' and Σ'', to take the limit $m \to \infty$ and then to consider the limit $h_\mu \to 0$. Once this selection of a particular μ has been made, we see from (21) that the horizontal effective coupling H gives rise to an external field $2H$, applied on Σ to the spin state μ, obtaining then the long-range order relative to the μ phase. So, we see that our parameter H has a role analogous to the effective field of the mean-field theory. By modulating through H the rigid boundary conditions on Σ' and Σ'', $\tilde{\psi}_0^{(1)}(\sigma)$ provides the simplest description of the phase having a temperature-dependent order parameter.

As we did for the disordered phase, we can improve $\tilde{\psi}_0^{(1)}(\sigma)$ through the action of $L^{\frac{1}{2}}$. So, in the ordered region, we have the sequence of one-parameter ansatz for $\psi_1(\sigma)$

$$\tilde{\psi}_0^{(1)}(\sigma), \tilde{\psi}_0^{(2)}(\sigma) = L^{\frac{1}{2}} \tilde{\psi}_0^{(1)}(\sigma), \ldots \tag{24}$$

and the relative RR quotients

$$\rho_0^{(\nu)}(H, K) = \frac{(\tilde{\psi}_0^{(1)}, L^\nu \tilde{\psi}_0^{(1)})}{(\tilde{\psi}_0^{(1)}, L^{\nu-1} \tilde{\psi}_0^{(1)})} \qquad (\nu \geqslant 1). \tag{25}$$

It is useful to point out that $\tilde{\psi}_{\rm o}^{(1)}(\sigma)$ can be written in the form

$$\tilde{\psi}_{\rm o}^{(1)}(\sigma) = \lim_{R\to+\infty} {\rm e}^{-(m-1)R} \sum_{\sigma'} {\rm e}^{R\sum_{i=1}^{m-1}\delta_{s'_i,s'_{i+1}}+H\sum_{i=1}^{m}\delta_{s'_i,s_i}+\frac{K}{2}\sum_{i=1}^{m-1}\delta_{s_i,s_{i+1}}}$$
$$\equiv \lim_{R\to+\infty} \tilde{\psi}_{\rm o}^{(1)}(\sigma; R) \tag{26}$$

and that

$$(\tilde{\psi}_{\rm o}^{(1)}, L^{\nu-1}\tilde{\psi}_{\rm o}^{(1)}) = Z_{\rm o}^{(\nu+2)}(H, K) \qquad (\nu \geqslant 1) \tag{27}$$

is the partition function of a Potts$(\nu+2; +\infty, H)$ model on a strip $\Sigma_{\nu+2}$ of m rows and $\nu+2$ columns, with the same coupling as the original model (1), except for the first and last column of sites, where we have a vertical NN coupling of infinite strength, and for the first and last column of rows where we have a horizontal NN coupling of strength H. This model provides an approximate description $\tilde{P}_{\rm o}^{(\nu)}$ of the RPD $P_\nu(\sigma_1, \sigma_2, \ldots, \sigma_\nu)$ in the ordered phase.

The Potts$(\nu+2; +\infty, H)$ is associated with a $q^{\nu+2} \times q^{\nu+2}$ transfer matrix $\ell_{\nu+2}(+\infty, H, K)$ having properties analogous to that of the previous $\ell = \ell_3(+\infty, H, K)$. If we call $\lambda_1^{(\nu+2)}(+\infty, H, K)$ its highest eigenvalue, we obtain the sequence of estimates $\phi_{\rm o}^{(\nu)}(K)$ of $\phi(K)$ in the form

$$\phi_{\rm o}^{(\nu)}(K) = \sup_{H\geqslant 0} \log \frac{\lambda_1^{(\nu+3)}(+\infty, H, K)}{\lambda_1^{(\nu+2)}(+\infty, H, K)} \equiv \sup_{H\geqslant 0} \phi_{\rm o}^{(\nu)}(H, K) \qquad (\nu = 1, 2, \ldots). \tag{28}$$

By taking a vector $\tilde{\psi}_{\rm d}^{(i)}(\sigma)$ of the sequence (15) and a vector $\tilde{\psi}_{\rm o}^{(j)}(\sigma)$ of the sequence (24), we obtain an ansatz

$$\tilde{\psi}(\sigma) = (\tilde{\psi}_{\rm d}^{(i)}(\sigma), \tilde{\psi}_{\rm o}^{(j)}(\sigma)) \tag{29}$$

with two *charts*, providing the estimates of $\phi(K)$, given by $\phi_{\rm d}^{(i)}(K)$ and $\phi_{\rm o}^{(j)}(K)$ respectively. These two functions are defined for any K. For each K, the choice between the two descriptions is made by a further application of the variational method, that is by considering

$$\max(\phi_{\rm d}^{(i)}(K), \phi_{\rm o}^{(j)}(K)). \tag{30}$$

We will say that two vectors $\tilde{\psi}_{\rm d}^{(i)}(\sigma)$ and $\tilde{\psi}_{\rm o}^{(j)}(\sigma)$ are exactly compatible if it happens that $\phi_{\rm d}^{(i)}(K)$ and $\phi_{\rm o}^{(j)}(K)$ match exactly at the transition point $K_{\rm t}$. In the next section, we will show that such vectors exist.

4. Duality

A crucial property of the 2D Potts model is the duality relation which connects the partition function in the high- and low-temperature regions. In terms of the free energy, it states that [2]

$$\phi(K) = \log x^2 + \phi(K^*) \tag{31}$$

where

$$x = \frac{{\rm e}^K - 1}{\sqrt{q}}$$

and K^*, the point *dual* of K, is defined by

$$\frac{{\rm e}^{K^*} - 1}{\sqrt{q}} = \frac{1}{x}.$$

The equation (31) is an exact analytic constraint, allowing the location of the transition point K_t (given by $x = 1$) on the basis of the non-analyticity of $\phi(K)$ at this point [2]. However, besides this formal aspect, we have no trace, in the above relation, of the mechanism leading to a transition from the disordered to the ordered phase.

Now, we will see that further insights and additional aspects of duality are provided by the variational method, by means of the ansatz's introduced in the previous section.

First of all we remark that, due to the fact that the ordered and disordered regions are described by a unique vector $\psi_1(\sigma)$ for finite m, the two *charts* of a proper ansatz should share some common property, as an expression of their common origin. As a matter of fact there is a property that allows us to associate univocally $\tilde{\psi}_o^{(i)}(\sigma)$ with $L^{\frac{1}{2}}\tilde{\psi}_d^{(i)}(\sigma) = \tilde{\psi}_d^{(i+1)}(\sigma)$, leading us to consider, as proper ansatz with two *charts*, the

$$\tilde{\psi}^{(1)}(\sigma) = (\tilde{\psi}_d^{(2)}(\sigma), \tilde{\psi}_o^{(1)}(\sigma)), \ldots, \tilde{\psi}^{(i)}(\sigma) = (\tilde{\psi}_d^{(i+1)}(\sigma), \tilde{\psi}_o^{(i)}(\sigma)) \tag{32}$$

and then the sequence of estimates $\phi^{(i)}(K)$ for $\phi(K)$, given by

$$\phi^{(i)}(K) = \max(\phi_d^{(i+1)}(K), \phi_o^{(i)}(K)) \qquad (i = 1, 2, \ldots). \tag{33}$$

This property, shared by $\tilde{\psi}_o^{(i)}(\sigma)$ and $\tilde{\psi}_d^{(i+1)}(\sigma)$, is an algebraic structure, as will be shown in the following.

We first consider $\tilde{\psi}_d^{(\nu)}(\sigma)$, for $\nu \geqslant 2$. We have seen that this vector defines the Potts(ν; A) model on a strip Σ_ν. Let us call $\tau = (t_1, t_2, \ldots, t_\nu)$ a spin configuration on a row of Σ_ν. The transfer matrix $\ell_\nu(A, K)$, which connects two adjacent rows, can be written in the form [2]

$$\ell_\nu(A, K) = V_\nu W_\nu \equiv q^{\frac{\nu}{2}} a^2 x^{\nu-2} \tilde{\ell}_\nu(A, K) \tag{34}$$

with

$$V_\nu = (I + xU_\nu^{(2)})(I + xU_\nu^{(4)}) \ldots (I + xU_\nu^{(2\nu-2)}) \qquad (\nu \geqslant 2)$$

$$W_\nu = \begin{cases} qa^2 \left(I + \dfrac{1}{a}U_2^{(1)}\right)\left(I + \dfrac{1}{a}U_2^{(3)}\right) & (\nu = 2) \\ q^{\frac{\nu}{2}} a^2 x^{\nu-2} \left(I + \dfrac{1}{a}U_\nu^{(1)}\right)\left(I + \dfrac{1}{a}U_\nu^{(2\nu-1)}\right)\left(I + \dfrac{1}{x}U_\nu^{(3)}\right) & \\ \qquad \ldots \left(I + \dfrac{1}{x}U_\nu^{(2\nu-3)}\right) & (\nu \geqslant 3) \end{cases} \tag{35}$$

where

$$a = \frac{e^{\frac{A+K}{2}} - 1}{\sqrt{q}}$$

and $U_\nu^{(1)}, U_\nu^{(2)}, \ldots, U_\nu^{(2\nu-1)}$ are $q^\nu \times q^\nu$ matrices defined by

$$U_\nu^{(2i-1)}(\tau, \tau') = \frac{1}{\sqrt{q}} \prod_{\substack{j=1 \\ j \neq i}} \delta_{t_j, t_j'} \qquad (i = 1, 2, \ldots, \nu)$$

$$U_\nu^{(2i)}(\tau, \tau') = \sqrt{q}\delta_{t_i, t_{i+1}} \prod_{j=1}^{\nu} \delta_{t_j, t_j'} \qquad (j = 1, 2, \ldots, \nu - 1).$$

They satisfy the relations

$$\begin{aligned}
&(U_\nu^{(i)})^2 = \sqrt{q}\,U_\nu^{(i)} \qquad (i = 1, 2, \ldots, 2\nu - 1)\\
&U_\nu^{(i)}U_\nu^{(j)} = U_\nu^{(j)}U_\nu^{(i)} \qquad (|i - j| \geqslant 2)\\
&U_\nu^{(i)}U_\nu^{(i+1)}U_\nu^{(i)} = U_\nu^{(i)} \qquad (i = 1, 2, \ldots, 2\nu - 2)\\
&U_\nu^{(i)}U_\nu^{(i-1)}U_\nu^{(i)} = U_\nu^{(i)} \qquad (i = 1, 2, \ldots, 2\nu - 1).
\end{aligned} \tag{36}$$

These relations define the Temperley–Lieb algebra generated by $U_\nu^{(1)}, U_\nu^{(2)}, \ldots, U_\nu^{(2\nu-1)}$, and then all the eigenvalues of $\ell_\nu(A, K)$, with the exception of their degeneracies.

Now we come to $\tilde{\psi}_{\mathrm{o}}^{(\nu)}(\sigma)(\nu = 1, 2, \ldots)$, which defines the Potts$(\nu+2; +\infty, H)$ model. It is useful to consider this model as the limit, for $R \to +\infty$, of Potts$(\nu+2; R, H)$, obtained through the substitution of $\tilde{\psi}_{\mathrm{o}}^{(1)}(\sigma)$ with $\tilde{\psi}_{\mathrm{o}}^{(1)}(\sigma; R)$ of equation (26). Then, following the previous procedure, we have that Potts$(\nu; R, H)(\nu \geqslant 3)$ is associated with a $q^\nu \times q^\nu$ transfer matrix $\ell_\nu(R, H, K)$, which can be written in the form

$$\ell_\nu(R, H, K) = \bar{V}_\nu \bar{W}_\nu(R) \tag{37}$$

with

$$\begin{aligned}
\bar{V}_\nu &= \begin{cases} (I + hU_\nu^{(2)})(I + hU_\nu^{(2\nu-2)})(I + xU_\nu^{(4)}) \ldots (I + xU_\nu^{(2\nu-4)}) & (\nu \geqslant 4)\\ (I + hU_3^{(2)})(I + hU_3^{(4)}) & (\nu = 3) \end{cases}\\
\bar{W}_\nu &= \mathrm{e}^{-2R} q^{\frac{\nu}{2}} \rho^2 x^{\nu-2} \left(I + \frac{1}{\rho}U_\nu^{(1)}\right)\left(I + \frac{1}{\rho}U_\nu^{(2\nu-1)}\right)\left(I + \frac{1}{x}U_\nu^{(3)}\right)\\
&\quad \ldots \left(I + \frac{1}{x}U_\nu^{(2\nu-3)}\right) \qquad (\nu \geqslant 3)
\end{aligned} \tag{38}$$

where

$$h = \frac{\mathrm{e}^H - 1}{\sqrt{q}} \qquad \rho = \frac{\mathrm{e}^R - 1}{\sqrt{q}}.$$

Taking the limit $R \to +\infty$, we obtain

$$\ell_\nu(+\infty, H, K) = \bar{V}_\nu \bar{W}_\nu(+\infty) \equiv q^{\frac{\nu-2}{2}} x^{\nu-2} \tilde{\ell}_\nu(+\infty, H, K) \tag{39}$$

with

$$\bar{W}_\nu(+\infty) = q^{\frac{\nu-2}{2}} x^{\nu-2} \left(I + \frac{1}{x}U_\nu^{(3)}\right) \ldots \left(I + \frac{1}{x}U_\nu^{(2\nu-3)}\right). \tag{40}$$

So, the matrices $U_\nu^{(1)}$ and $U_\nu^{(2\nu-1)}$ being suppressed by the limiting procedure, we remain only with the $2(\nu - 1) - 1$ $q^\nu \times q^\nu$ matrices $U_\nu^{(2)}, U_\nu^{(3)}, \ldots, U_\nu^{(2\nu-3)}, U_\nu^{(2\nu-2)}$. However, if we define

$$\bar{U}_\nu^{(i)} = U_\nu^{(i+1)} \qquad (i = 1, 2, \ldots, 2(\nu - 1) - 1)$$

we see from (36), that these $2(\nu - 1) - 1$ $q^\nu \times q^\nu$ matrices $\bar{U}_\nu^{(i)}$ satisfy the same algebraic relations which are satisfied by the $q^{\nu-1} \times q^{\nu-1}$ matrices $U_{\nu-1}^{(i)}$ $(i = 1, 2, \ldots, 2(\nu - 1) - 1)$. We only have with the $\bar{U}_\nu^{(i)}$ a representation of different dimensionality of the same algebraic structure. Together with the structure of $\tilde{\ell}_\nu(+\infty, H, K)$ and $\tilde{\ell}_{\nu-1}(A, K)$, this is the common property which allows us to couple the vector $\tilde{\psi}_{\mathrm{o}}^{(\nu-2)}(\sigma)(\nu \geqslant 3)$ with the vector $\tilde{\psi}_{\mathrm{d}}^{(\nu-1)}(\sigma)$, leading then to the sequence (32).

As a matter of fact, it follows from (34), (35), (38)–(40) that the eigenvalues of $\tilde{\ell}_\nu(+\infty, H, K)$ coincide with those of $\tilde{\ell}_{\nu-1}(A_D(H, K^*), K^*)$, where $A_D(H, K^*)$ is defined by

$$\frac{\mathrm{e}^{\frac{A_D(H,K^*)+K^*}{2}} - 1}{\sqrt{q}} = \frac{1}{h}. \tag{41}$$

In particular, we have

$$\lambda_1^{(\nu)}(+\infty, H, K) = \frac{h^2}{\sqrt{q}} x^{2\nu-5} \lambda_1^{(\nu-1)}(A_D(H, K^*), K^*) \qquad (\nu \geqslant 3) \tag{42}$$

which is a duality relation for the partition function of Potts(ν; $+\infty$, H) and Potts($\nu-1$; A).

As a consequence, coming back to the variational method, we obtain the duality relation for the RR quotients

$$\frac{\lambda_1^{(i+3)}(+\infty, H, K)}{\lambda_1^{(i+2)}(+\infty, H, K)} = x^2 \frac{\lambda_1^{(i+2)}(A_D(H, K^*), K^*)}{\lambda_1^{(i+1)}(A_D(H, K^*), K^*)}. \tag{43}$$

Since, as H goes from zero to $+\infty$, $A_D(H, K^*)$ goes from $+\infty$ to $-K^*$, we deduce from (43), (19), (28) that

$$\phi_{\mathrm{o}}^{(i)}(K) = \log x^2 + \phi_{\mathrm{d}}^{(i+1)}(K^*)$$

or, equivalently,

$$\phi_{\mathrm{d}}^{(i+1)}(K) = \log x^2 + \phi_{\mathrm{o}}^{(i)}(K^*) \tag{44}$$

the relative sup being attained at *dual* points, according to (41). So $\tilde{\psi}_{\mathrm{o}}^{(i)}(\sigma)$ and $\tilde{\psi}_{\mathrm{d}}^{(i+1)}(\sigma)$ are exactly compatible.

The equation (44) is the variational version of (31). Indeed, from (44) it follows that

$$\phi_{\mathrm{o}}^{(i)}(K) - \phi_{\mathrm{d}}^{(i+1)}(K) = \phi_{\mathrm{d}}^{(i+1)}(K^*) - \phi_{\mathrm{o}}^{(i)}(K^*). \tag{45}$$

Therefore, if at a point K we have, for example, $\phi_{\mathrm{d}}^{(i+1)}(K) > \phi_{\mathrm{o}}^{(i)}(K)$, then at the dual point it results that $\phi_{\mathrm{o}}^{(i)}(K^*) > \phi_{\mathrm{d}}^{(i+1)}(K^*)$. So we conclude that the estimates $\phi^{(i)}(K)$ given by (33), satisfy exactly the duality relation (31), that is

$$\phi^{(i)}(K) = \log x^2 + \phi^{(i)}(K^*) \qquad (i = 1, 2, \ldots). \tag{46}$$

We see that the mechanism of non-analyticity at K_{t} can be described as a crossing phenomenon, involving $\phi_{\mathrm{o}}^{(i)}(K)$ and $\phi_{\mathrm{d}}^{(i+1)}(K)$.

5. Two-parameters ansatz

The above results can be extended to two-parameters ansatz, which are useful to consider both from the computational and variational point of view.

To fix our ideas, let us suppose that we need to go beyond the estimates $\phi_{\mathrm{d}}^{(2)}(K)$ and $\phi_{\mathrm{d}}^{(3)}(K)$. We can consider then $\phi_{\mathrm{d}}^{(4)}(K)$. The relative calculations require strips of four and five columns, to be compared with the case of three and four columns needed for $\phi_{\mathrm{d}}^{(3)}(K)$. However, we can also improve $\phi_{\mathrm{d}}^{(3)}(K)$ without changing the number of columns and the size of the involved transfer matrices, by adding a further parameter. In fact, by making use of the procedure of an approximate description of the averaging process, we can introduce the two-parameters ansatz for $\psi_1(\sigma)$

$$\tilde{\psi}_{\mathrm{d}}^{\prime(3)}(\sigma) = \left(\sum_{\sigma'} \mathrm{e}^{A \sum_{i=1}^{m-1} \delta_{s_i', s_{i+1}'} + B \sum_{i=1}^{m} \delta_{s_i', s_i}} \right) \mathrm{e}^{\frac{K}{2} \sum_{i=1}^{m-1} \delta_{s_i, s_{i+1}}} \tag{47}$$

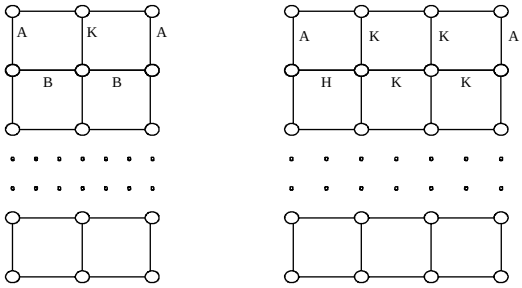

Figure 1. The three- and four-column Potts models originated by the two-parameters ansatz (47). They correspond to the partition functions $Z_{\rm d}^{(3)}$ and $Z_{\rm d}^{(4)}$ leading to our estimate $\phi_{\rm d}^{'(3)}$ of the free energy in the disordered phase.

with $A, B \geqslant 0$.

Then, by considering $(\tilde{\psi}_{\rm d}^{'(3)}, L\tilde{\psi}_{\rm d}^{'(3)})/(\tilde{\psi}_{\rm d}^{'(3)}, \tilde{\psi}_{\rm d}^{'(3)})$, we are led to the estimate of $\phi(K)$

$$\phi_{\rm d}^{'(3)}(K) = \sup_{A,B} \lim_{m\to\infty} \frac{1}{m} \log \frac{Z_{\rm d}^{(4)}(A, B, K)}{Z_{\rm d}^{(3)}(A, B, K)} \tag{48}$$

where $Z_{\rm d}^{(i)}(A, B, K)(i = 3, 4)$ are the partition functions of the Potts models shown in figure 1. Therefore, we deduce from (19) and (48)

$$\phi_{\rm d}^{(3)}(K) \leqslant \phi_{\rm d}^{'(3)}(K) \leqslant \phi(K). \tag{49}$$

We can improve analogously $\phi_{\rm o}^{(2)}(K)$, which is determined through $\tilde{\psi}_{\rm o}^{(2)}(\sigma) = L^{\frac{1}{2}}\tilde{\psi}_{\rm o}^{(1)}(\sigma)$, by introducing the two-parameter ansatz

$$\tilde{\psi}_{\rm o}^{'(2)}(\sigma) = L^{\frac{1}{2}}\tilde{\psi}_{\rm o}^{'(1)}(\sigma) \tag{50}$$

with

$$\tilde{\psi}_{\rm o}^{'(1)}(\sigma) = \left(\sum_{\sigma'} \Pi(\sigma', \sigma)\right) {\rm e}^{\frac{G}{2}\sum_{i=1}^{m-1}\delta_{s_i,s_{i+1}}} \qquad (G \geqslant 0) \tag{51}$$

where $\Pi(\sigma', \sigma)$ is given by (21). Through $(\tilde{\psi}_{\rm o}^{'(2)}, L\tilde{\psi}_{\rm o}^{'(2)})/(\tilde{\psi}_{\rm o}^{'(2)}, \tilde{\psi}_{\rm o}^{'(2)})$, we obtain the estimate

$$\phi_{\rm o}^{'(2)}(K) = \sup_{G,H} \lim_{m\to\infty} \frac{1}{m} \log \frac{Z_{\rm o}^{(5)}(G, H, K)}{Z_{\rm o}^{(4)}(G, H, K)} \tag{52}$$

where $Z_{\rm o}^{(i)}(G, H, K)$ $(i = 4, 5)$ are the partition functions of the Potts models shown in figure 2.

By making use again of the generators of the Temperley–Lieb algebra we have, as in (44),

$$\phi_{\rm d}^{'(3)}(K) = \log x^2 + \phi_{\rm o}^{'(2)}(K^*) \tag{53}$$

so that the estimate of $\phi(K)$,

$$\phi^{'(2)}(K) = \max(\phi_{\rm d}^{'(3)}(K), \phi_{\rm o}^{'(2)}(K)) \tag{54}$$

also satisfies the duality relation (31). As a matter of fact, the parameters H and G are the *dual* expression of A and B, respectively.

In principle, further improved two-parameters ansatz can be obtained from the action of $(L^{\frac{1}{2}})^i$ on the vector $\tilde{\psi}^{'(2)}(\sigma) = (\tilde{\psi}_{\rm d}^{'(3)}(\sigma), \tilde{\psi}_{\rm o}^{'(2)}(\sigma))$, leading to the sequence of estimates $\phi^{'(i+2)}(K)$.

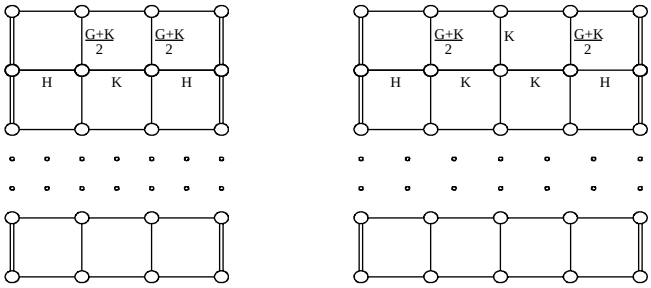

Figure 2. The four- and five-column Potts models originated by the two-parameters ansatz (50). They correspond to the partition functions $Z_{\mathrm{o}}^{(4)}$ and $Z_{\mathrm{o}}^{(5)}$ leading to our estimate $\phi_{\mathrm{o}}^{\prime(2)}$ of the free energy in the ordered phase. Double lines connect spins constrained to be in the same state.

6. Some numerical results and comments

Now we discuss the effective aspects of the variational method presented in the previous sections, by considering the estimates of lower order.

We have calculated explicitly $\phi_{\mathrm{d}}^{(2)}(K)$, $\phi_{\mathrm{d}}^{(3)}(K)$, $\phi_{\mathrm{d}}^{\prime(3)}(K)$, $\phi_{\mathrm{o}}^{(1)}(K)$, $\phi_{\mathrm{o}}^{(2)}(K)$ and $\phi_{\mathrm{o}}^{\prime(2)}(K)$, obtaining then $\phi^{(1)}(K)$, $\phi^{(2)}(K)$ and $\phi^{\prime(2)}(K)$, according to (33) and (54). We have also considered $\phi_{\mathrm{d}}^{(1)}(A, K)$, which does not have a *dual* partner in our scheme, but, in any case, allows us to obtain the simplest approximation to $\phi(K)$, for every K. As a matter of fact, $\phi_{\mathrm{d}}^{(1)}(A, K)$ has a finite absolute maximum point $\bar{A}(K)$ for $K < \bar{K}_{\mathrm{t}}$, while for $K > \bar{K}_{\mathrm{t}}$ the absolute maximum is taken at $A \to +\infty$. These points are also relative maximum points in a neighbourhood of $\bar{K}_{\mathrm{t}}$, at its right and at its left, respectively. $\bar{K}_{\mathrm{t}}$ is an estimate of the transition point K_{t}; at $\bar{K}_{\mathrm{t}}$ we have a crossing between $\phi_{\mathrm{d}}^{(1)}(\bar{A}, K)$ and $\phi_{\mathrm{d}}^{(1)}(+\infty, K)$.

The simplest estimate of our variational method is then given by

$$\phi^{(0)}(K) = \max(\phi_{\mathrm{d}}^{(1)}(\bar{A}(K), K), \phi_{\mathrm{d}}^{(1)}(+\infty, K)). \tag{55}$$

It turns out that $\phi^{(0)}(K)$ satisfies quite well, even if approximately, the duality relation (31), when q is large. As a matter of fact, by making a comparison with the known exact results, we have that $\phi^{(0)}(K)$ provides quite a reliable description, for every K, of the properties of the 2D square Potts model for $q \geqslant 50$, the description being more and more accurate as q increases. In this respect, our simplest approximation behaves like mean-field theory. However, from a quantitative point of view, we get a net improvement, since it turns out that the mean-field estimates, which are exact in the limit $q \to +\infty$, differ significantly from the known exact results already at q of order 10^3. Indeed, in the large q expansions, we have power series in $1/\sqrt{q}$ with very large coefficients [9], which are responsible for such deviations. In figures 3(*a*) and (*b*) we plot, for $q = 100$, $\phi^{(0)}(K)$ and the relative internal energy $u^{(0)}(K)$, given by

$$\frac{u^{(0)}(K)}{-J} = \frac{\mathrm{d}}{\mathrm{d}K}\phi^{(0)}(K)$$

and we make a comparison with the analogous mean-field quantities $\phi_{\mathrm{MF}}(K)$ and $u_{\mathrm{MF}}(K)$; the small circles are the known exact values.

From a computational point of view, in order to avoid $q^{\nu} \times q^{\nu}$ matrices of very high order, especially if q is large, we have made use of the relationship between the reduced transfer matrices and the matrices $U_{\nu}^{(1)}, \ldots, U_{\nu}^{(2\nu-1)}$ which satisfy the relations (36) (see (34), (35), (37) and (38)). Now the relations (36) are also satisfied by the $4^{\nu} \times 4^{\nu}$ matrices

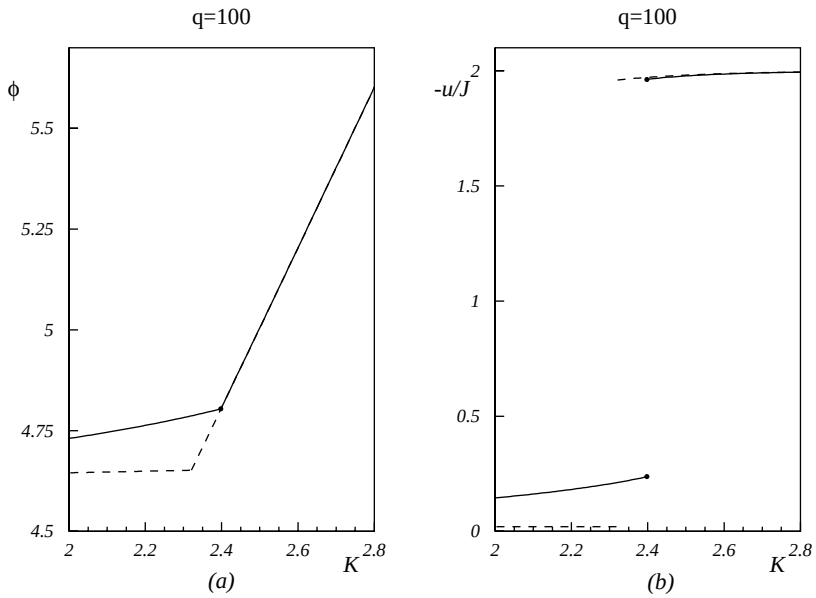

Figure 3. Our simplest estimate (full lines) of the free energy (*a*) and of the internal energy (*b*) at $q = 100$ are compared with the results from the mean-field theory (broken lines). Also shown are the exact values at the transition point (small circles).

$U_\nu^{(i)}$ given by [2]

$$U_\nu^{(i)}(\alpha_1, \alpha_2, \ldots, \alpha_{2\nu}|\alpha'_1, \alpha'_2, \ldots, \alpha'_{2\nu}) = \delta_{\alpha_1\alpha'_1} \ldots \delta_{\alpha_{i-1}\alpha'_{i-1}} h_{\alpha_i\alpha_{i+1}} h_{\alpha'_i\alpha'_{i+1}} \delta_{\alpha_{i+2}\alpha'_{i+2}} \ldots \delta_{\alpha_{2\nu}\alpha'_{2\nu}} \tag{56}$$

$(i = 1, 2, \ldots, 2\nu - 1)$, where $\alpha_i = \pm 1$, $\alpha'_i = \pm 1$,

$$h_{++} = h_{--} = 0 \qquad h_{+-} = \mathrm{e}^{-\frac{\lambda}{2}} \qquad h_{-+} = \mathrm{e}^{\frac{\lambda}{2}}$$

with $2\cosh\lambda = q^{1/2}$. This is the 6-vertex representation of the Temperley–Lieb algebra.

The eigenvalues of our reduced transfer matrices have been determined by making use of the representation (56), which allows us to translate a Potts strip Σ_i into an ice-type (or 6-vertex) strip Σ'_i and, as a by-product, to consider also non integer values of q.

As expected, the results of our calculations show that K_t is the only crossing point of $\phi_\mathrm{d}^{(i+1)}(K)$ with $\phi_\mathrm{o}^{(i)}(K)(i = 1, 2)$ and also of $\phi_\mathrm{d}^{'(3)}(K)$ with $\phi_\mathrm{o}^{'(2)}(K)$. Furthermore $\phi_\mathrm{d}^{(i+1)}(K) > \phi_\mathrm{o}^{(i)}(K)$, $\phi_\mathrm{d}^{'(3)}(K) > \phi_\mathrm{o}^{'(2)}(K)$ for $K < K_\mathrm{t}$, so that

$$\phi^{(i)}(K) = \begin{cases} \phi_\mathrm{d}^{(i+1)}(K) & K \leqslant K_\mathrm{t} \\ \phi_\mathrm{o}^{(i)}(K) & K \geqslant K_\mathrm{t} \end{cases} \qquad \phi^{'(2)}(K) = \begin{cases} \phi_\mathrm{d}^{'(3)}(K) & K \leqslant K_\mathrm{t} \\ \phi_\mathrm{o}^{'(2)}(K) & K \geqslant K_\mathrm{t}. \end{cases} \tag{57}$$

In order to see the degree of accuracy which we obtain, in table 1 we compare our results at K_t, for several q, with the known exact values $\phi(K_\mathrm{t})$ and $-u_\mathrm{d}(K_\mathrm{t})/J$. Of course, due to the duality relation, we reproduce exactly, through $\phi^{(1)}(K)$, $\phi^{(2)}(K)$ and $\phi^{'(2)}(K)$, the mean value $(u_\mathrm{d}(K_\mathrm{t}) + u_\mathrm{o}(K_\mathrm{t}))/2$.

We see that, for every $q > 4$, $\phi^{(0)}(K_\mathrm{t})$, $\phi^{(1)}(K_\mathrm{t})$, $\phi^{(2)}(K_\mathrm{t})$ and $\phi^{'(2)}(K_\mathrm{t})$ provide a monotonous sequence of very accurate estimates of $\phi(K_\mathrm{t})$, the accuracy being better as q increases. At $q = 5$, the errors of these estimates are of 0.25%, 0.10%, 0.05%, 0.02%

Table 1. Our predictions for the free energy ϕ, the internal energy in the disorder phase u_d and the specific heat in the ordered phase C_o calculated at the transition point K_t. The exact values for the free internal energy and the internal energy are also reported.

q	5	10	15	20	30	100
$\phi^{(0)}(K_t)$	2.403 806	2.895 964	3.202 524	3.428 507	3.758 386	4.803 843
$\phi^{(1)}(K_t)$	2.407 461	2.898 352	3.204 045	3.429 543	3.758 951	4.803 915
$\phi^{(2)}(K_t)$	2.408 681	2.899 057	3.204 432	3.429 775	3.759 055	4.803 922
$\phi^{'(2)}(K_t)$	2.409 378	2.899 363	3.204 563	3.429 841	3.759 078	4.803 923
$\phi^{''(2)}(K_t)$	2.410 306	2.899 597	3.204 631	3.429 867	3.759 084	4.803 923
$\phi(K_t)$	2.409 849	2.899 522	3.204 615	3.429 862	3.759 083	4.803 923
$-u_d^{(0)}(K_t)/J$	1.189 595	0.868 464	0.701 564	0.597 935	0.474 256	0.236 934
$-u_d^{(1)}(K_t)/J$	1.249 544	0.911 736	0.728 555	0.615 741	0.483 472	0.237 961
$-u_d^{(2)}(K_t)/J$	1.282 729	0.933 513	0.739 933	0.622 203	0.486 151	0.238 115
$-u_d^{'(2)}(K_t)/J$	1.317 667	0.949 882	0.746 210	0.625 067	0.487 035	0.238 139
$-u_d^{''(2)}(K_t)/J$	1.413 784	0.971 819	0.751 328	0.626 773	0.487 391	0.238 142
$-u_d(K_t)/J$	1.420 754	0.968 203	0.750 492	0.626 529	0.487 353	0.238 142
$C_o^{(1)}(K_t)$	4.504 9	5.406 3	4.602 4	3.864 1	2.905 9	1.176 2
$C_o^{(2)}(K_t)$	6.235 0	7.292 7	5.695 6	4.514 4	3.192 3	1.196 9
$C_o^{'(2)}(K_t)$	9.942 1	9.852 6	6.692 8	4.970 0	3.335 0	1.201 3
$C_o^{''(2)}(K_t)$	30.632 4	15.700 9	7.957 2	5.372 8	3.416 5	1.202 2

respectively, while at $q = 10$, the sequence of the errors is 0.12%, 0.04%, 0.02%, 0.005%. As a matter of fact, by analysing the stability of our results for $K \neq K_t$, we obtain that these errors are still rapidly decreasing, as we move from K_t.

For the internal energy, our predictions are characterized by larger errors, provided q is low and K is very near to K_t. But our last estimate, through the two-parameters variational ansatz, is quite accurate for every K and every $q \gtrsim 15$. At $q = 20$, $-u_d^{'(2)}(K_t)/J$ gives the exact value with an error of 0.1%. On the other hand, we obtain errors of 7% and 2% at $q = 5$ and $q = 10$, respectively. However, as we move from K_t, our estimates of the internal energy are again quite accurate, even for low q's. This is shown in figures 4(a) and (b), where we plot the functions $-u^{(1)}(K)/J$, $-u^{(2)}(K)/J$ and $-u^{'(2)}(K)/J$.

Through $\phi^{(i)}(K)$ ($i = 0, 1, 2$) and $\phi^{'(2)}(K)$ we have also determined the estimates $C^{(i)}(K)$ and $C^{'(2)}(K)$ of the specific heat $C(K)$, which is not exactly known, even at K_t. The sequence $C^{(0)}(K), C^{(1)}(K), C^{(2)}(K), C^{'(2)}(K)$ is monotonous and rapidly converging for every K, provided $q \gtrsim 30$. For these q's, $C^{'(2)}(K)$ provides a quite reliable estimate of $C(K)$ for every K. This is also true for $C^{(0)}(K), C^{(1)}(K)$ and $C^{(2)}(K)$ as soon as q increases. For lower values of q, we have to move from K_t in order to get the same accuracy. In figure 5 we plot $C^{(1)}(K), C^{(2)}(K)$ and $C^{'(2)}(K)$ at $q = 10$, while in table 1 we give, for several q, $C_o^{(1)}(K_t), C_o^{(2)}(K_t)$ and $C_o^{'(2)}(K_t)$, that is the previous estimates at K_t, evaluated in the ordered phase (the predictions for the disordered phase can be obtained from these through duality).

On the other hand, we obtain a considerable improvement in our estimates of the internal energy and of the specific heat for lower q and K very near to K_t, by accelerating the convergence of the already rapidly convergent sequence of our estimates of the free energy. Starting from $\phi^{(1)}(K)$, $\phi^{(2)}(K)$ and $\phi^{'(2)}(K)$ and by making use of the Aitken Δ^2

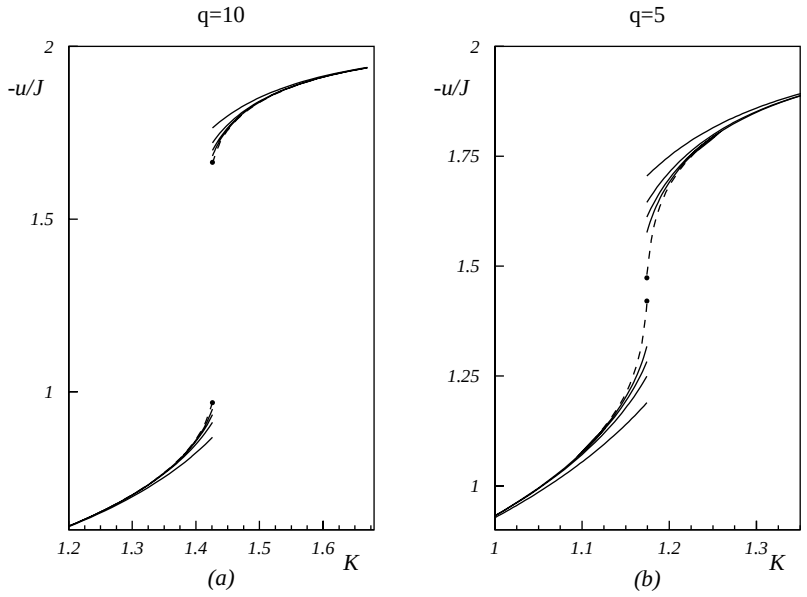

Figure 4. The monotonous sequence of our estimates $-u^{(1)}(K)/J, -u^{(2)}(K)/J$ and $-u^{'(2)}(K)/J$ of the internal energy at (*a*) $q = 10$ and (*b*) $q = 5$. The broken curve represents $-u^{''(2)}(K)/J$, obtained from equation (58). Also shown are the exact values at the transition point (small circles).

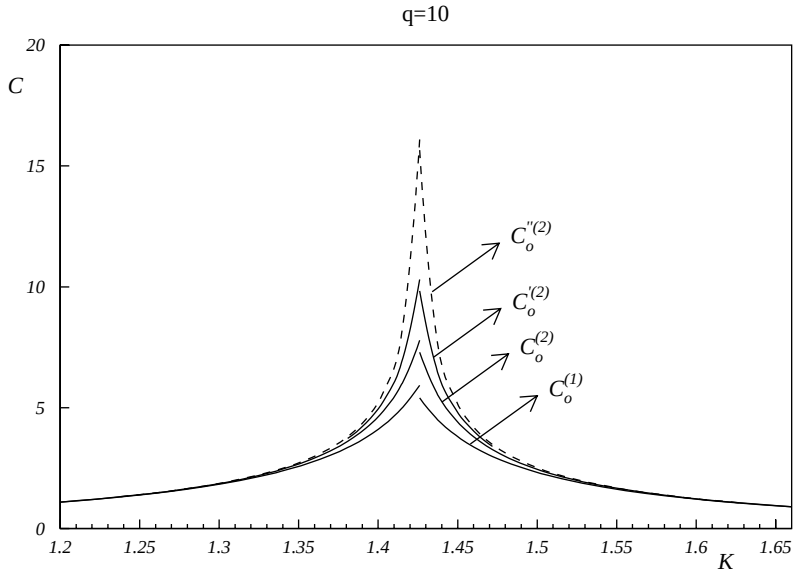

Figure 5. The monotonous sequence of our estimates $C^{(1)}(K), C^{(2)}(K)$ and $C^{'(2)}(K)$ of the specific heat at $q = 10$. The broken curve represents $C^{''(2)}(K)$, obtained from equation (58).

method (or of the [1, 1] Padè approximant), we deduce the new extimate

$$\phi^{''(2)}(K) = \phi^{(1)}(K) + \frac{\phi^{(2)}(K) - \phi^{(1)}(K)}{1 - \frac{\phi^{'(2)}(K) - \phi^{(2)}(K)}{\phi^{(2)}(K) - \phi^{(1)}(K)}}. \tag{58}$$

We call $u^{''(2)}(K)$ and $C^{''(2)}(K)$ the internal energy and the specific heat calculated through $\phi^{''(2)}(K)$. The improvement which we obtain can be checked by comparing $u_{\mathrm{d}}^{''(2)}(K_{\mathrm{t}})$ with the known exact value. We give in table 1 $\phi^{''(2)}(K_{\mathrm{t}})$, $u_{\mathrm{d}}^{''(2)}(K_{\mathrm{t}})$ and $C_{\mathrm{o}}^{''(2)}(K_{\mathrm{t}})$ for several q. We see that $-u_{\mathrm{d}}^{''(2)}(K_{\mathrm{t}})/J$ reproduces $-u_{\mathrm{d}}(K_{\mathrm{t}})/J$ within the errors of 0.5% and 0.4% at $q = 5$ and $q = 10$, respectively, a net improvement with respect to $-u_{\mathrm{d}}^{'(2)}(K_{\mathrm{t}})/J$. On the other hand, when $u_{\mathrm{d}}^{'(2)}(K)$ and $C^{'(2)}(K)$ are already quite reliable, $\phi^{''(2)}(K)$ does not lead to significant corrections. For $q = 10$ and $q = 5$, the function $-u_{\mathrm{d}}^{''(2)}(K)/J$ is plotted in figures 4(*a*) and (*b*), where its role is made quite evident. A stronger correction is obtained for specific heat, as shown in figure 5, where we plot $C^{''(2)}(K)$, for $q = 10$. Our estimate of 16 for $C_{\mathrm{o}}(K_{\mathrm{t}})$ at $q = 10$, agrees essentially with the estimate of 18 obtained from large-q expansions [9], which is contradicted by the value 32 deduced from low-temperature series [6]. For $q = 15, 20, 30$ the results of [9] are confirmed quite well by our $C_{\mathrm{o}}^{''(2)}(K_{\mathrm{t}})$.

So, we see that the variational approach presented in this paper, besides providing exact predictions and physical or mathematical mechanisms for the first-order transition of our model (for $q > 4$), is also very effective in the numerical predictions of unknown quantities.

We point out that the first approximation in the Baxter method [22] corresponds to the variational approximation of Kramers–Wannier (K–W) type [20, 21]. On the other hand the K–W approximation corresponds to the form of the cluster variation method introduced by Kikuchi [26], which improves both the mean-field and the Bethe approximations.

The K–W type of ansatz is improved by Baxter [22] in a way formally elegant, but not very flexible for the purpose of taking into account some general non-perturbative aspects.

Our sequence $\phi^{(1)}(K), \phi^{(2)}(K), \ldots$ is an improvement of $\phi^{(0)}(K)$ (55), which also coincides exactly with the K–W approximation in its disordered phase. However, our $\phi^{(1)}(K), \phi^{(2)}(K), \ldots$ are all able to localize exactly the point of singularity of the thermodynamical quantities and to get exactly other properties related to duality. Such a feature is to be compared with the approximate (in any case quite good) localization of the critical point in the Baxter procedure [23, 24].

Our numerical analysis will be completed in a future paper and will concern properties such as spontaneous magnetization and susceptibility, which are accessible through low-temperature series expansions [6], but not through large-q expansions. On the other hand, since these two approaches give contradicting results in the case of specific heat, a further analysis of the estimates made in [6] would be appropriate.

Furthermore we will explore more completely the consequences of duality, which in the variational approach proposed here appears at a more fundamental level. Through the reduced probability distributions and the variational parameters having dual partners, it is possible to study more completely the link between typical properties of the ordered and disordered phases.

Our aim will also be to extend the above approach to the important case of the 3D Potts model.

Acknowledgments

This research was supported by Ministero dell'Università e della Ricerca Scientifica (40% and 60% funds) and Istituto Nazionale di Fisica Nucleare.

References

[1] Wu F Y 1982 *Rev. Mod. Phys.* **54** 235
[2] Baxter R J 1982 *Exactly Solved Models in Statistical Mechanics* (New York: Academic)
[3] Buffenoir E and Wallon S 1993 *J. Phys. A: Math. Gen.* **26** 3045
[4] Klümper A, Schadschneider A and Zittarz J 1989 *Z. Phys.* B **76** 247
[5] Binder K 1987 *Rep. Prog. Phys.* **50** 783
[6] Briggs K M, Enting I G and Guttmann A J 1994 *J. Phys. A: Math. Gen.* **27** 1503
[7] Bhattacharya T, Lacaze R and Morel A 1993 *Europhys. Lett.* **23** 547
[8] Bhattacharya T, Lacaze R and Morel A 1994 *Nucl. Phys. B (Proc. Suppl.)* **34** 671
[9] Bhattacharya T, Lacaze R and Morel A 1997 *J. Physique* I **7** 81
[10] Billoire A 1992 *Nucl. Phys. B (Proc. Suppl.)* **26** 589
[11] Billoire A, Neuhaus T and Berg B 1993 *Nucl. Phys.* B **396** 779
[12] Janke W and Kappler S 1994 *Nucl. Phys. B (Proc. Suppl.)* **34** 674
[13] Janke W and Kappler S 1995 *Europhys. Lett.* **31** 345
[14] Janke W and Kappler S 1995 *Nucl. Phys. B (Proc. Suppl.)* **42** 770
[15] Mittag L and Stephen M J 1974 *J. Phys. A: Math. Gen.* **7L** 109
[16] Pearce P A and Griffiths R B 1980 *J. Phys. A: Math. Gen.* **13** 2143
[17] Huang K 1987 *Statistical Mechanics* (New York: Wiley)
[18] Morita T 1972 *J. Math. Phys.* **13** 115
[19] Camp W J and Fisher M E 1972 *Phys. Rev.* B **6** 946
[20] Munster A 1974 *Statistical Thermodynamics* vol 2 (Berlin: Springer)
[21] Rujan P 1979 *Physica* **96A** 379
[22] Baxter R J 1978 *J. Stat. Phys.* **19** 461
[23] Kelland S B 1976 *Can. J. Phys.* **54** 1621
[24] Tsang S K 1979 *J. Stat. Phys.* **20** 95
[25] Savit R 1980 *Rev. Mod. Phys.* **52** 453
[26] Burley D M 1972 Closed form approximations for lattice systems *Phase Transitions and Critical Phenomena* vol 2, ed C Domb and M S Green (New York: Academic)